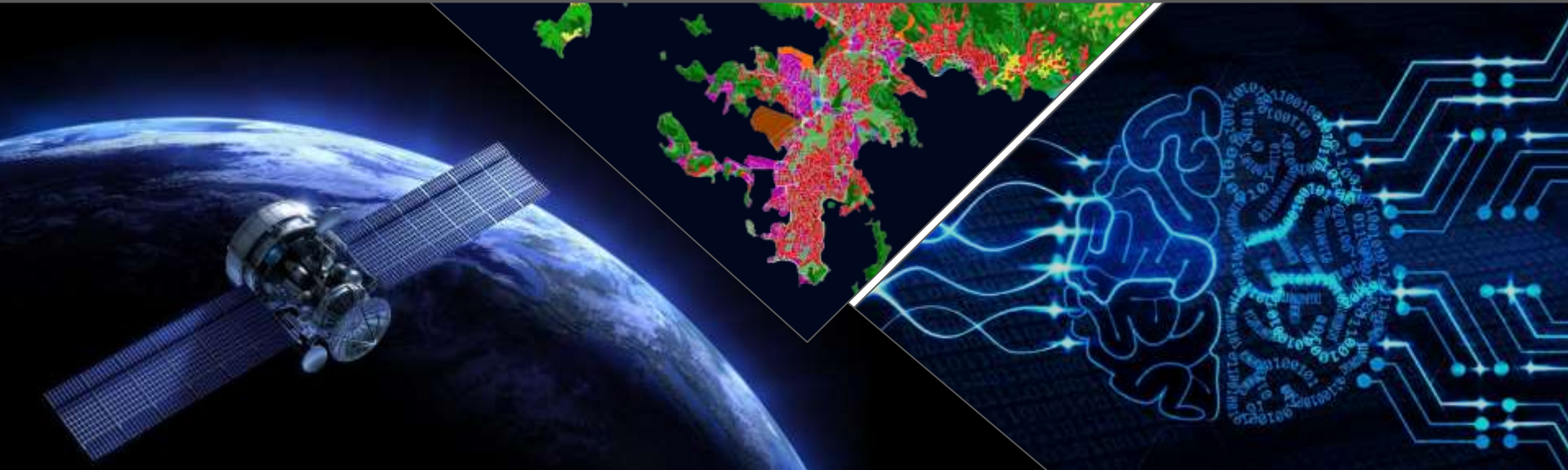


# Données satellitaires et Deep Learning : Application à l'occupation du sol de Nouvelle-Calédonie

Guillaume Rousset, Morgan Mangeas, Marc Despinoy, Dominique Simpelaere



- 
- ❖ Accueil dans l'UMR Espace-DEV Nouméa
    - Directeurs de thèse
      - *Morgan Mangeas* (Espace-DEV)
      - *Dominique Simpelaere* (UNC)
    - Encadrant
      - *Marc Despinoy* (Espace-DEV)
  
  - ❖ Deuxième année de thèse
  
  - ❖ Collaborations
    - ETH-Zürich avec *Konrad Shindler*
    - Projets TOSCA (Apureza et Babaçu)
    - Observatoire de l'Environnement de Nouvelle-Calédonie (données)



**ETH** zürich

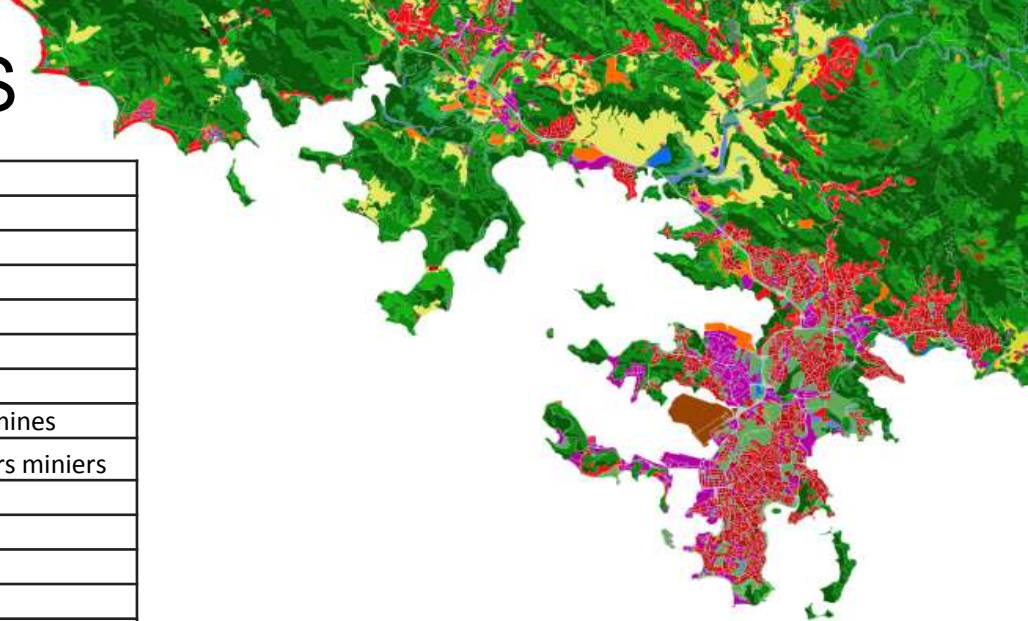


# Le contexte de la Nouvelle-Calédonie

- ❖ Une île haute tropicale avec de nombreux type de sols
  - forêt sèche, forêt tropical, maquis, ...
- ❖ Un enjeu environnemental important
  - Une flore, une faune, et un lagon unique
  - Exploitation minière importante
  - Des impacts climatiques forts
- ❖ Le Mode d'Occupation des Sols (MOS)
  - Projet initié par l'ŒIL
  - Un produit en 2010 et 2014
  - La Province Sud, 7000 km<sup>2</sup> de surface
  - Précision géographique de 1 ha, pour une échelle d'utilisation de 1/15,000 à 1/18,000



# La nomenclature du MOS



L1	L2	L3	Description	
1			Territoires artificialisés	
	11		Zones urbanisées	
		111	Tissu urbain continu	
		112	Tissu urbain discontinu	
		113	Habitat isolé	
	12		Zones industrielles ou commerciales et équipements	
	13		Décharges, chantiers, extraction de matériaux : hors mines	
	14		Mines, décharges minières, infrastructures et chantiers miniers	
15		Espaces verts artificialisés et équipements		
16		Réseaux de communication		
2			Territoires agricoles	
	21		Terres arables et cultures permanentes	
	23		Terres pastorales aménagées structurées	
	24		Plantation d'arbres – sylviculture	
3			Formation végétale	
	31		Strate arborée	
		311	Strate arborée	
	32		Strate arbustive et herbacée	
		321	Strate arbustive	
		322	Strate herbacée	
	33			Espaces ouverts, sans ou avec peu de végétation
		331		Plages, dunes et sable
		332		Roches et sols nus
		333		Végétation clairsemée
334			Zones incendiées	

L1	L2	L3	Description
4			Zones humides
	41		Zones humides maritimes
		411	Marais maritimes
42		Zones humides intérieures	
5			Surfaces en eau
	51		Eaux continentales
	52		Eaux maritimes

# Les données satellites utilisées

## ❖ SPOT6

- Résolution : 1,5 mètres
- Bandes : RGB et PIR
- MOS 2014

## ❖ Sentinel-2

- Résolution : 10 mètres
- Bande : RGB, red-edge et PIR
- Travail en cours

## ❖ RapidEye

- Résolution : 5 mètres
- Bande : RGB, red-edge et PIR
- MOS 2010
- Travail en cours



# Changement de la nomenclature du MOS

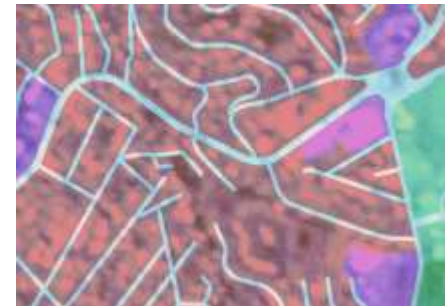
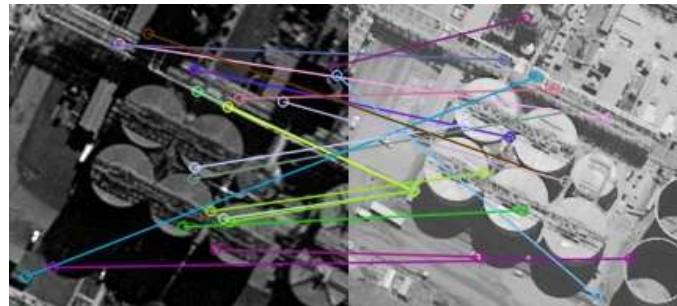
## ❖ Problèmes liés aux données satellites

- léger décalage
- résolution de l'image
- nuages, ombres

## ❖ Limitations du « machine learning »

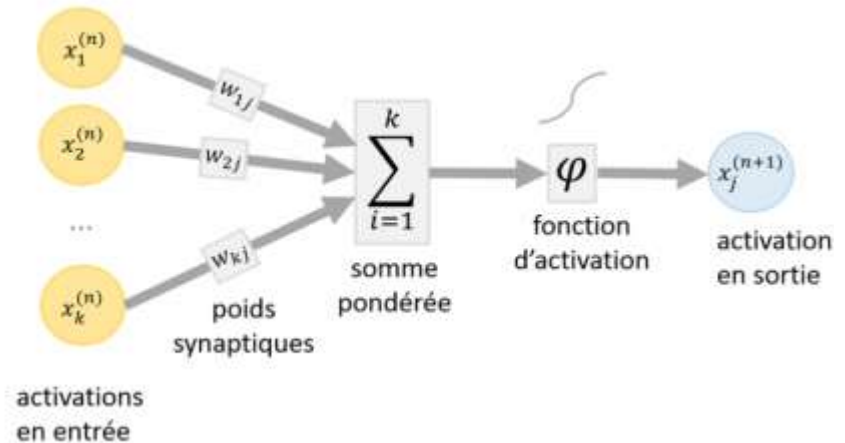
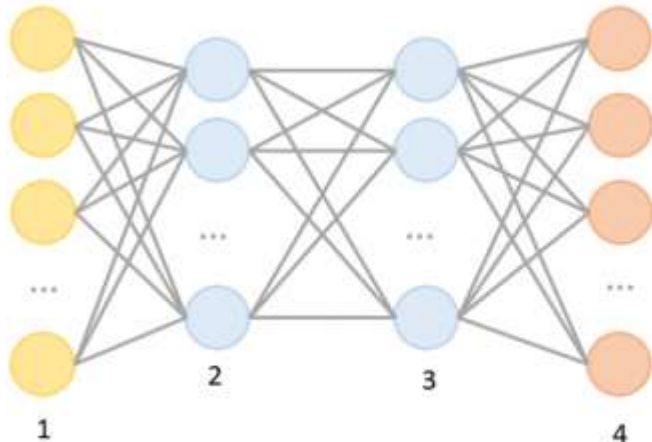
- Données exogènes
- Précisions trop grandes
- Changement de la nomenclature

L1	L2	Description
1		Territoires artificialisés
	11	Zones urbanisées
	12	Zones industrielles ou commerciales et équipements
	13	Décharges, chantiers, extraction de matériaux
	14	Réseaux de communication
2		Formation naturelle et agriculture
	21	Strate arborée
	22	Strate arbustive
	23	Espaces ouverts, sans ou avec peu de végétation
	24	Roches
	25	Zones incendiées
3		Zones humides
4		Surfaces en eau



# Le Deep Learning ?

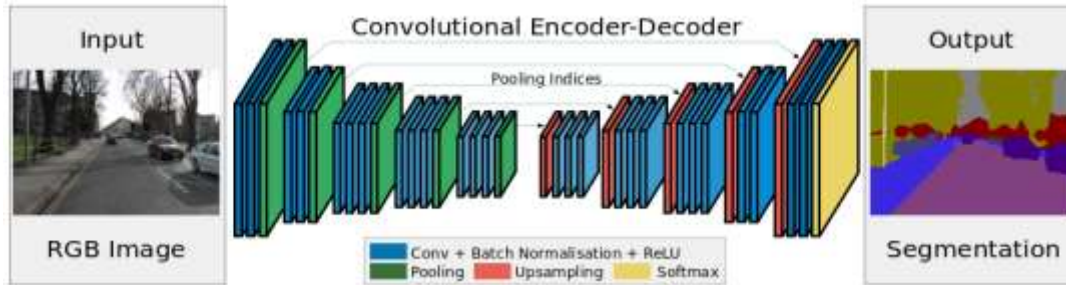
- ❖ Basé sur les réseaux de neurones
- ❖ Une des techniques les plus performantes dans le domaine de la reconnaissance d'objets
- ❖ Un taux de reconnaissance proche des performances humaines



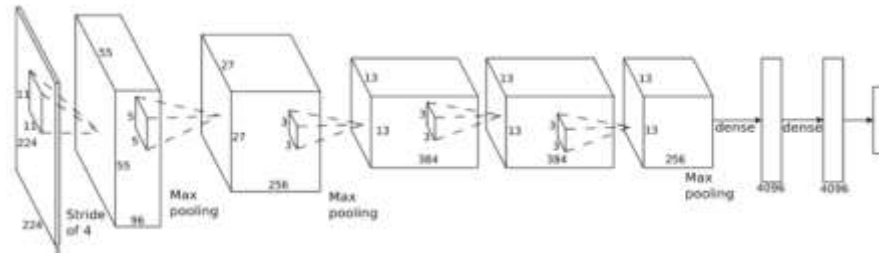
# Les modèles utilisés

## ❖ Deux techniques de Deep Learning pour le moment

### ➤ DeepNetForEO



### ➤ AlexNet : une architecture simple

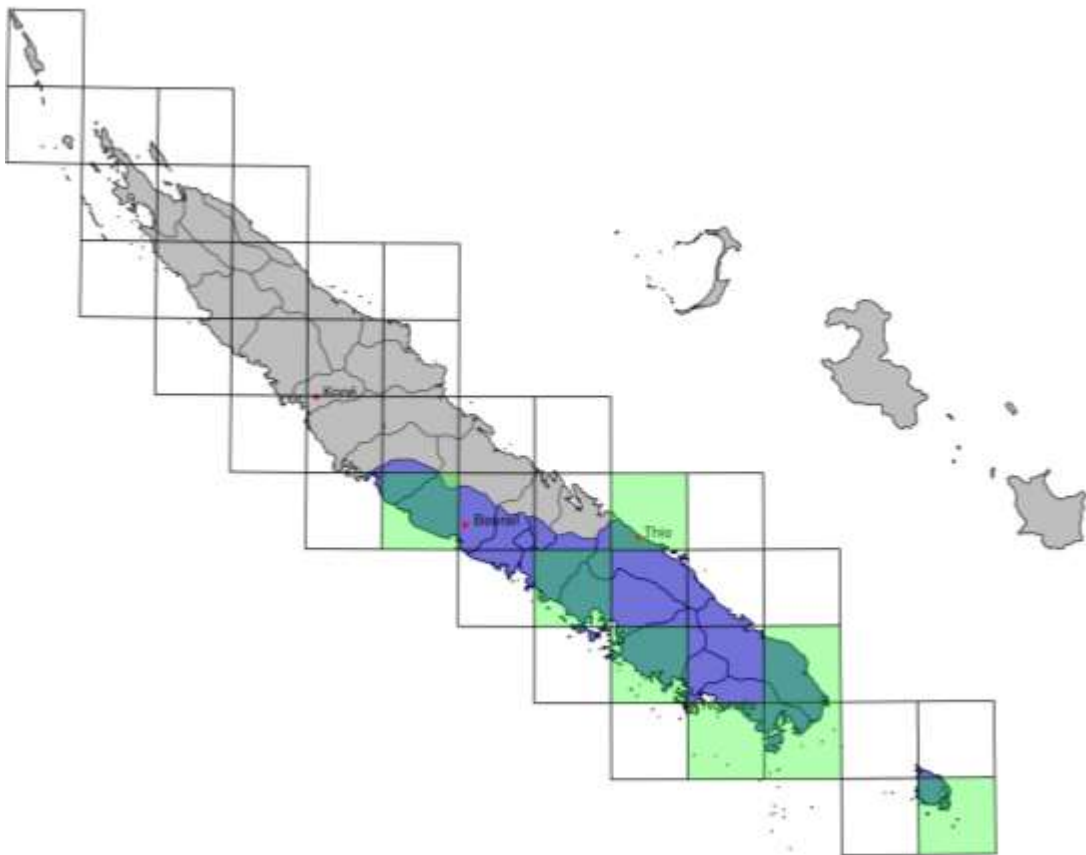


## ❖ Une technique de machine learning plus classique

### ➤ XGBoost



# La sélection des données



- ❖ Découpage en fenêtre 64x64
- ❖ Fenêtre glissante pour la création d'un index
  - 64/4
  - 64/32

		classes					
		index	1	2	3	...	m
nb pixels de la classe inclus dans l'imagette	0	x	x	x	...	x	
	1	0	x	x	...	x	
	2	0	x	0	...	0	
	3	x	x	x	...	x	
	...	...	...	...	...	...	
	n-2	x	x	x	...	x	
	n-1	0	x	x	...	x	
n	x	x	0	...	0		
<b>total</b>	n+1	n1	n2	n3	...	n5	

		classes					
		index	1	2	4	...	m
nb pixels de la classe inclus dans	0	x	x	x	...	x	
	2	0	x	0	...	0	
	...	...	...	...	...	...	
	n-2	x	x	0	...	x	
	n	x	x	0	...	0	
	<b>total</b>	n+1-1000	n1'	n2'	n3'	...	n4'

# Les difficultés du MOS

- ❖ Problèmes liés au MOS
  - finesse des classes à l'intérieur
  - Mauvais choix de classe (oubli, confusion)



# Résultat sur le jeu de test

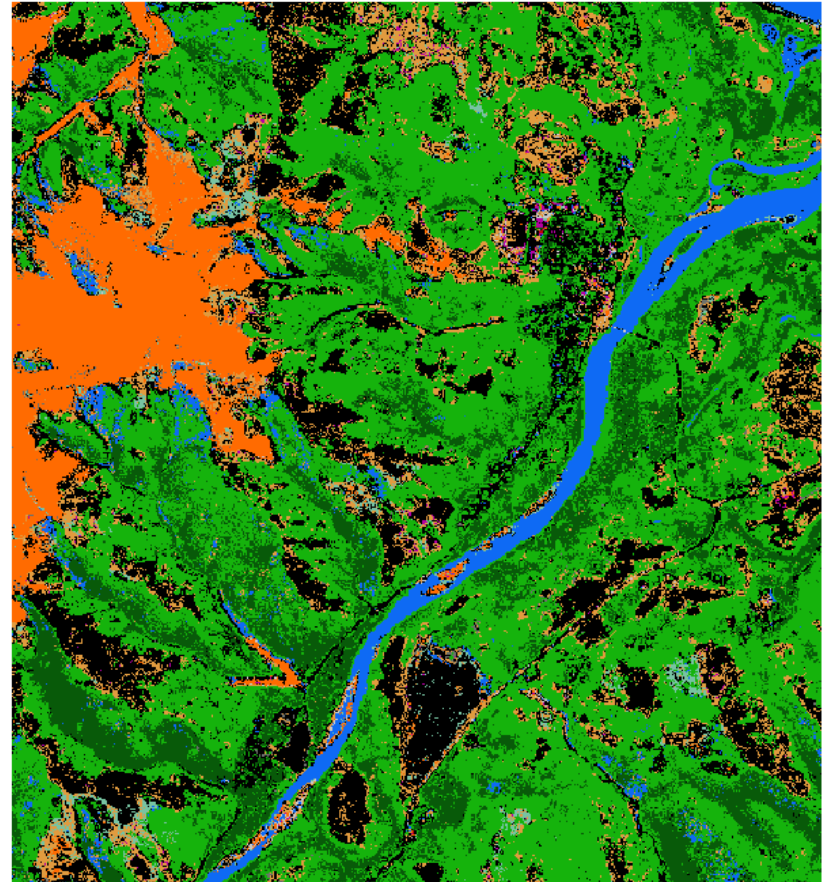
Classes	XGBoost	DeepNetForEO	AlexNet
Zones urbanisées	5,93	<b>70,40</b>	68,20
Zones industrielles	17,00	<b>44,37</b>	36,30
Mines et chantiers	59,11	<b>81,81</b>	78,02
Réseaux de communication	4,73	3,78	<b>44,86</b>
Strate arborée	64,81	66,22	<b>71,42</b>
Strate arbustive	<b>70,85</b>	69,95	58,86
Espaces ouverts	43,07	65,19	<b>73,42</b>
Roches	8,01	28,51	<b>51,67</b>
Zones incendiées	37,01	<b>51,60</b>	47,08
Zones humides	17,17	<b>51,71</b>	18,16
Surface en eau	50,52	<b>54,11</b>	53,94
Précision globale	39,92	58,76	<b>60,19</b>

# Résultats visuels de ces architectures



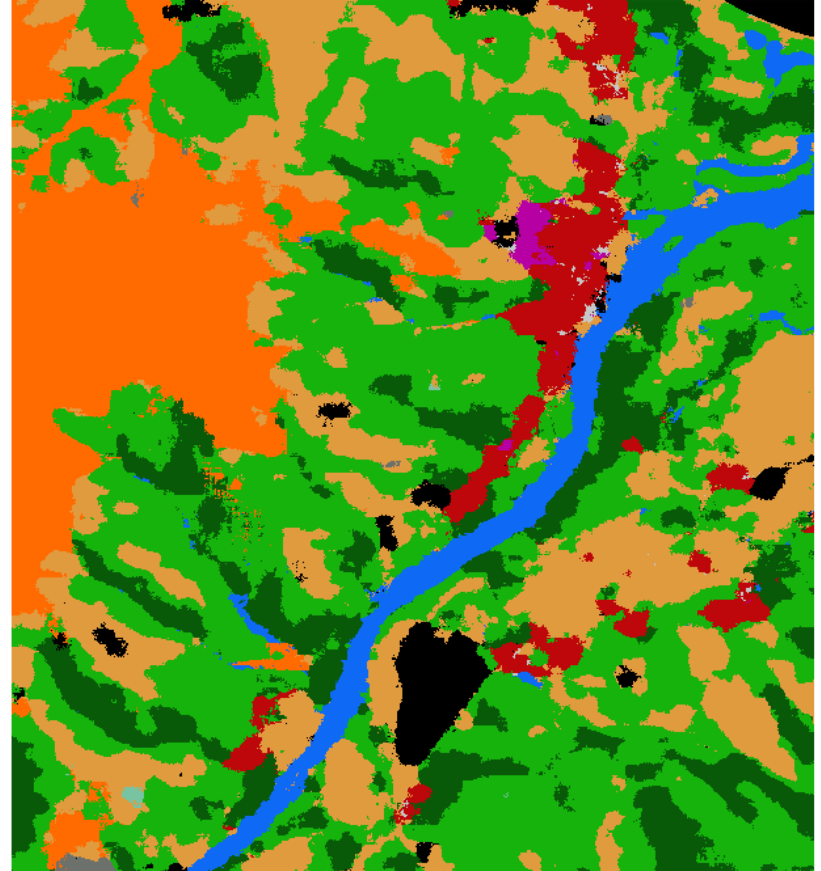


# Résultats visuels de ces architectures

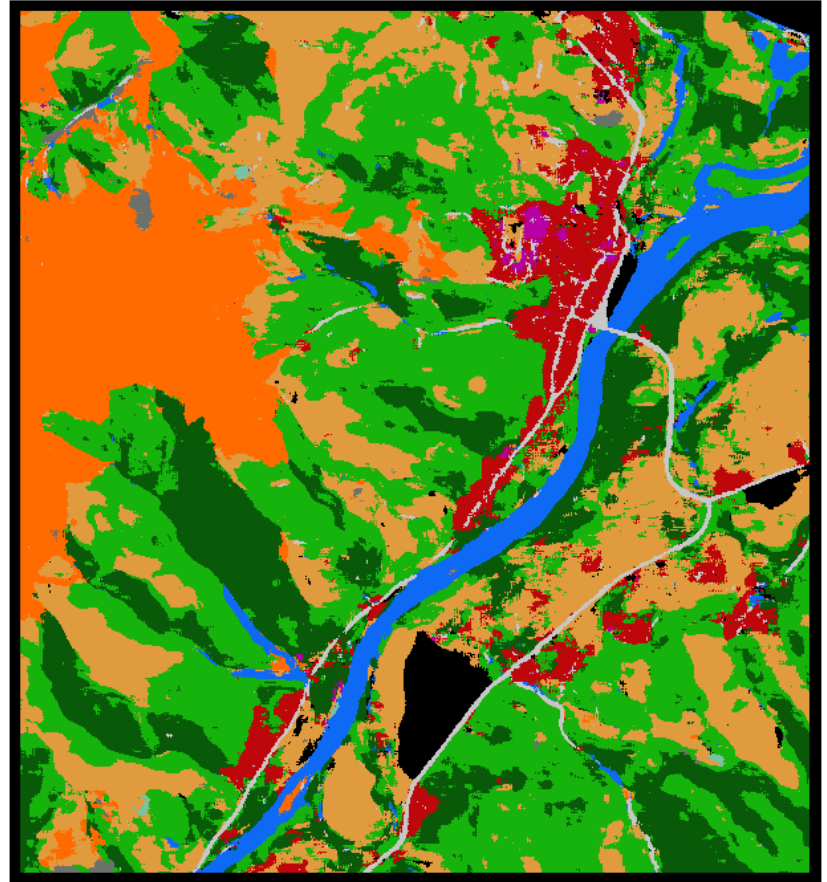




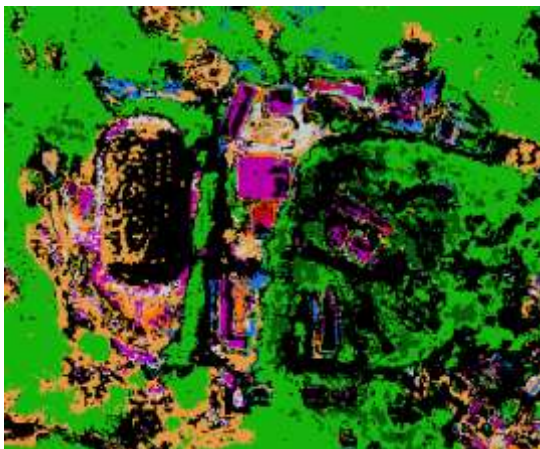
# Résultats visuels (DeepNetForEO)



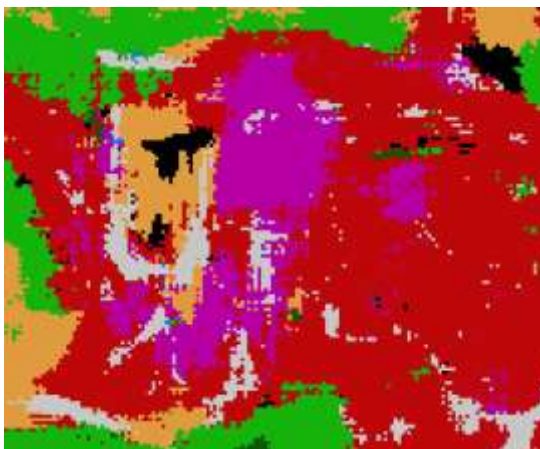
# Résultats visuels (AlexNet)



# Résultats visuels

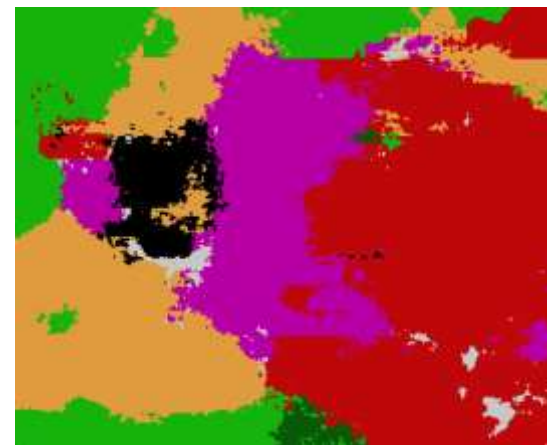


XGBoost



AlexNet

DeepNetForEO



# Conclusion

## Constitution du jeu de données

- ❖ Pré-traitement
- ❖ Découpage de la scène
- ❖ Création du jeu de données

## Création du modèle

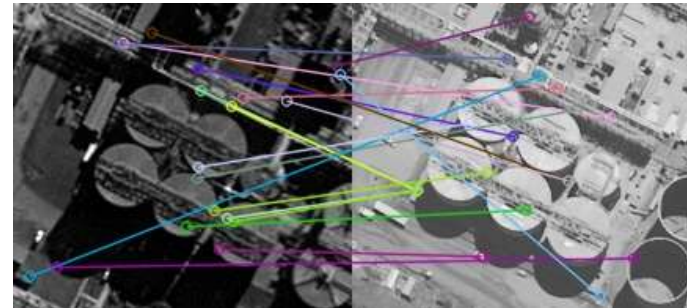
- ❖ Type de modèle
- ❖ Architecture
- ❖ Paramètres

## Apprentissage et résultats

- ❖ Visualisation
- ❖ Sortie du modèle
- ❖ Décision

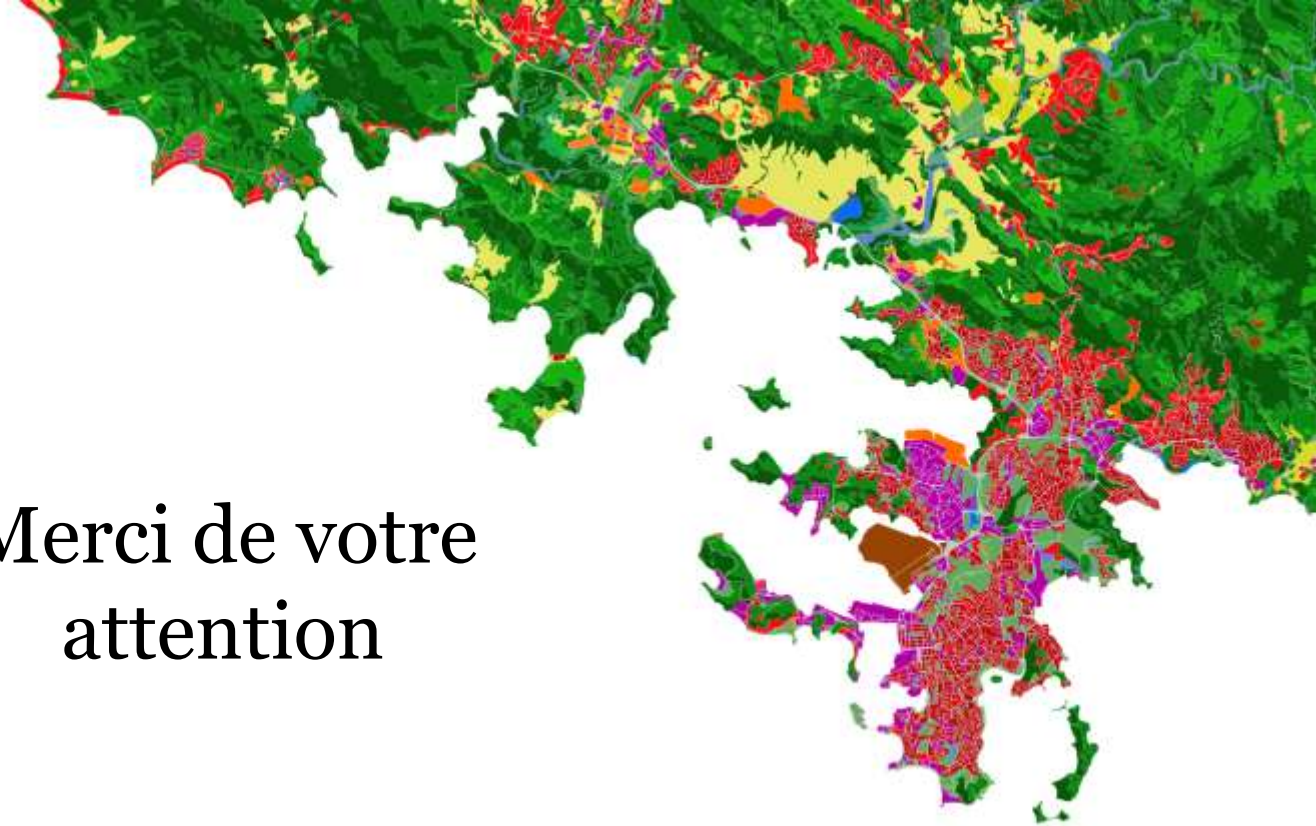
# Perspectives

- ❖ Application d'autres méthodes de machine learning et d'autres architectures de deep learning
- ❖ « Negative mining » pour améliorer les résultats
- ❖ Applications de méthodes statistiques à la fin pour lisser les classes
- ❖ Amélioration de l'architecture (architecture pyramidale => multi-résolution)
- ❖ Meilleur calage des données à l'aide d'une architecture « Deep Learning » pour les données SPOT6.  
Inspiration du SIFT





Merci de votre  
attention



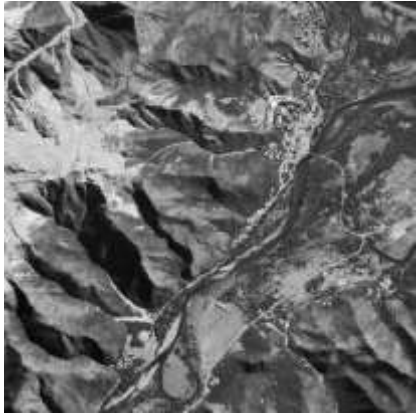
# Méthode de référence : indices+textures+xgboost

## ❖ Xgboost

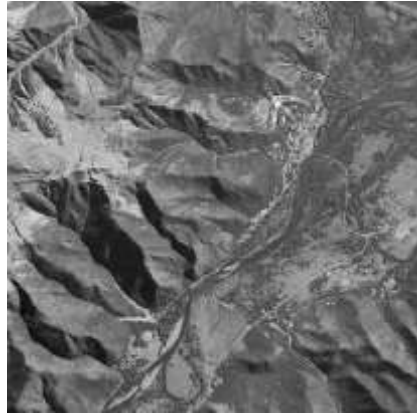
- Ensemble de modèles de prédiction dits « faibles »
- Basé sur des arbres de décision
- Calibration en optimisant une fonction de coût dérivable par une méthode d'optimisation du type gradient
- Très performant, très rapide et adapté au traitement de données de gros volume
- Créé par Tianqi Chen (Université de Washington)
- Packages R, python, C++, Java...

# Canaux bruts

R



G



B

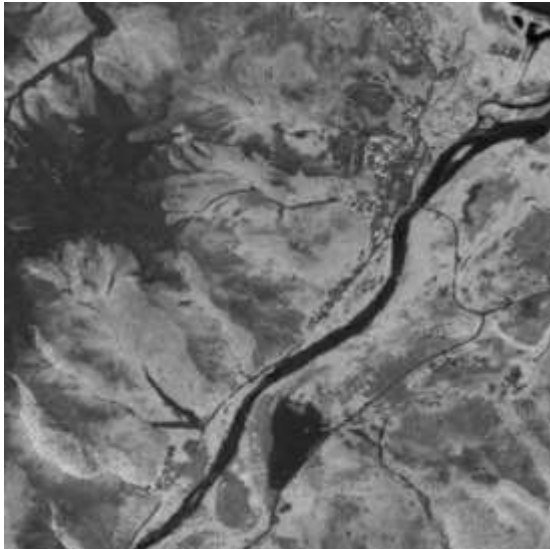


PIR

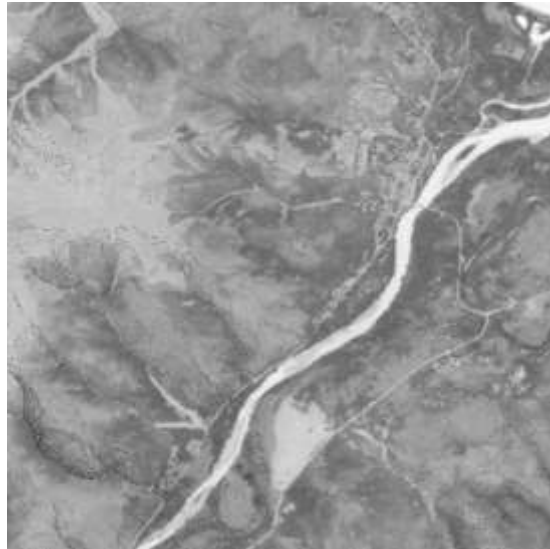


# Indices

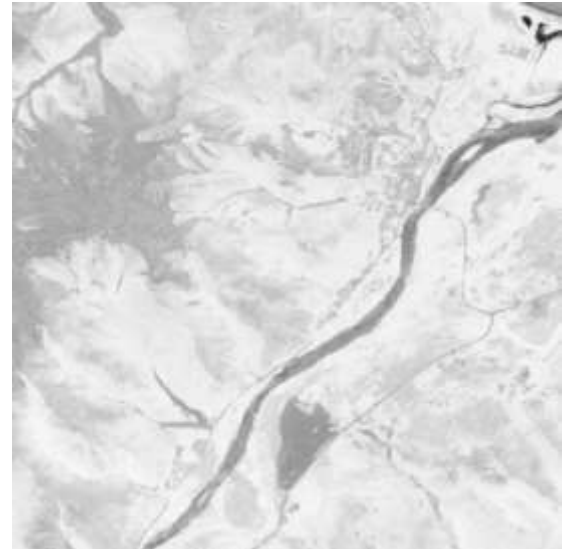
NDVI



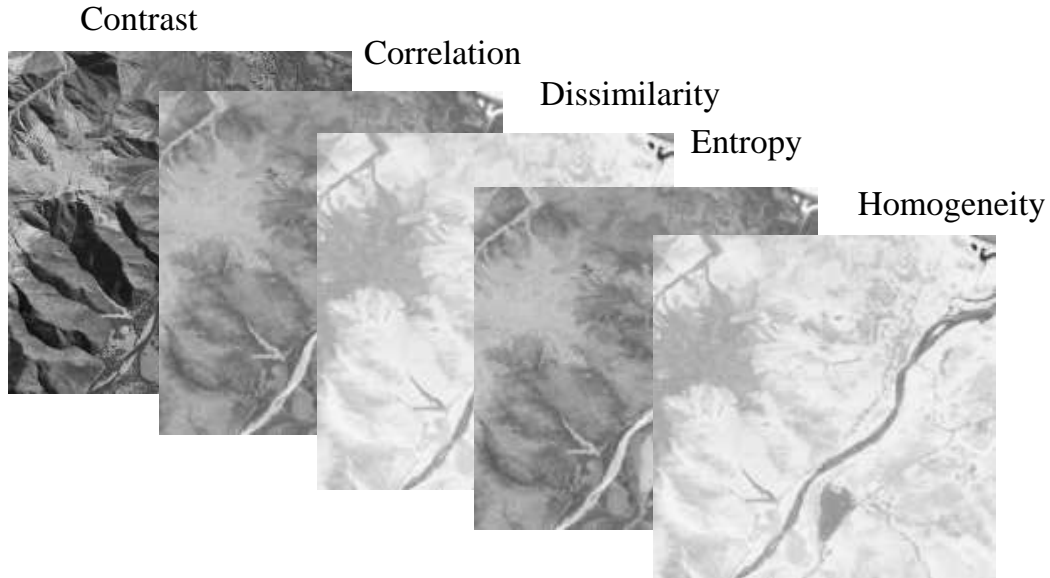
MNDWI



MSAVI



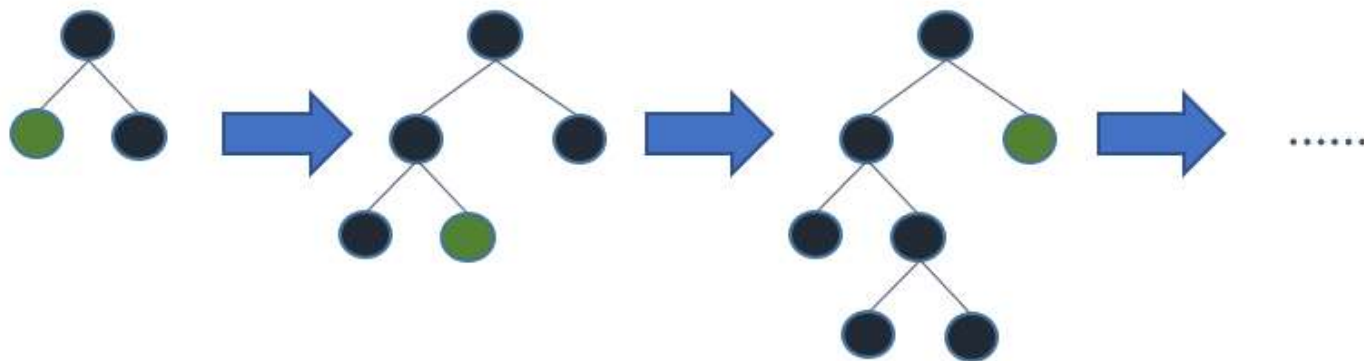
# Textures



Pour chaque canal  
Pour chaque indice



# Xgboost



Leaf-wise tree growth